Robust Multisource Remote Sensing Image Registration Method Based on Scene Shape Similarity

Ming Hao, Jian Jin, Mengchao Zhou, Yi Tian, and Wenzhong Shi

Abstract

Image registration is an indispensable component of remote sensing applications, such as disaster monitoring, change detection, and classification. Grayscale differences and geometric distortions often occur among multisource images due to their different imaging mechanisms, thus making it difficult to acquire feature points and match corresponding points. This article proposes a scene shape similarity feature (SSSF) descriptor based on scene shape features and shape context algorithms. A new similarity measure called $SSSF_{ncc}$ is then defined by computing the normalized correlation coefficient of the SSSF descriptors between multisource remote sensing images. Furthermore, the tie points between the reference and the sensed image are extracted via a template matching strategy. A global consistency check method is then used to remove the mismatched tie points. Finally, a piecewise linear transform model is selected to rectify the remote sensing image. The proposed $SSSF_{ncc}$ aims to extract the scene shape similarity between multisource images. The accuracy of the proposed $SSSF_{ncc}$ is evaluated using five pairs of experimental images from optical, synthetic aperture radar, and map data. Registration results demonstrate that the $SSSF_{ncc}$ similarity measure is robust enough for complex nonlinear grayscale differences among multisource remote sensing images. The proposed method achieves more reliable registration outcomes compared with other popular methods.

Introduction

Remote sensing applications require the interactive use of multisource remote sensing images due to the diversity of remote sensing image acquisition methods. Multisource remote sensing images usually reflect different characteristics of ground features and complement surface monitoring. These images must be registered in geospatial settings to integrate the required information. Image registration, which is a basic preprocessing step in remote sensing data processing, is the process of superimposing images acquired under different time phases, different sensors, and different shooting conditions (Zitová and Flusser 2003). Its accuracy has an essential influence on subsequent image processing, such as detection change, object identification, and image fusion (Hao *et al.* 2014). Although automatic image registration techniques have greatly progressed, selecting tie points manually is often necessary when these techniques are applied to multisource image registration due to time and geometric differences and

local distortion between multispectral images (Goncalves *et al.* 2011). Grayscale differences among multisource images make it more challenging to detect tie points rather than among linear single-spectral images. This study addresses this issue by defining an accurate matching strategy that is robust for multispectral images with nonlinear grayscale differences.

In accordance with the image registration process, remote sensing image registration methods can be divided into feature-based and area-based methods (Zitová and Flusser 2003). Feature-based methods must first detect and then describe the features from a reference image. Afterward, the tie points are extracted using the similarity of these features. Common features consist of point (Yu *et al.* 2008), line (Sui *et al.* 2015), and region features (Goncalves *et al.* 2011). Recently, local feature descriptors have also been applied to image registration. The scale-invariant feature transform (SIFT) (Lowe 2004) is the most representative local feature descriptor. This descriptor has been successfully used for remote sensing image registration due to its invariance to geometric differences and image scales (Ma *et al.* 2010; Sedaghat *et al.* 2015). Accordingly, many improved SIFT algorithms, such as the C-SIFT (Abdel-Hakim *et al.* 2006), have been proposed. However, these algorithms are not effective for optical and synthetic aperture radar (SAR) images with significant nonlinear grayscale differences (Kelman *et al.* 2007; Ye *et al.* 2014). These feature-based methods can adapt well to geometric differences between images. When large nonlinear grayscale differences exist between multisource images, extracting common features from multisource images robustly is difficult. The repetition rate of feature point detection is greatly reduced, thus resulting in poor image registration (Suri and Reinartz 2010). Therefore, these methods are not reliable for multisource remote sensing image registration.

Area-based methods, which are sometimes called "template matching", are another kind of image registration. These methods first define a template window of a certain size on the sensed image and then uses a similarity measure as a basis to find a corresponding template area on the reference image. The method then selects the center point of the template as the tie point. Finally, the optimal geometric transformation relationship between images is determined in accordance with the tie points. As mentioned above, similarity measures have a vital influence on area-based methods. Commonly used similarity measures mainly consist of the sum of squared differences (SSD), normalized correlation coefficient (NCC), and mutual information (MI). Among these similarity measures, the SSD is probably the simplest similarity measure in image

Ming Hao, Jian Jin, Mengchao Zhou, and Yi Tian are with NASG Key Laboratory of Land Environment and Disaster Monitoring, China University of Mining and Technology, Xuzhou, China (1715882901@qq.com).

Wenzhong Shi is with Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China (lswzshi@polyu.edu.hk).

registration that can reflect the similarity of the grayscale differences between two images; the smaller the SSD, the more similar the two images are. Although the SSD has high computational efficiency, it cannot accurately adapt to multispectral remote sensing images. The NCC is a classical similarity measure in image registration. It has been widely used for single-spectral image registration due to its linear invariance for grayscale changes and high computational efficiency (Inglada *et al.* 2004). However, it performs poorly in nonlinear radiometric differences (Hel-Or *et al.* 2014). By contrast, the MI is more resistant to complex nonlinear grayscale changes and is gradually used for multisource image registration (Brunner *et al.* 2010; Cole-Rhodes *et al.* 2003; Suri and Reinartz 2010). Unfortunately, MI-based image matching methods have numerous calculations, thereby limiting its application. In summary, these similarity measures are not well suited for handling radiometric distortions among multisource images, because they are more stable for intensity changes. Therefore, previous researchers have applied these metrics to image descriptors, such as wavelet-like (Hong *et al.* 2008; Murphy *et al.* 2016) and gradient features (Ravanbakhsh *et al.* 2013), to improve registration performance. However, these features cannot reflect the common properties of multisource images.

Recently, shape features have been integrated into similarity descriptors for image matching and received better registration results compared with conventional similarity measures in multisource image processing (Fu *et al.* 2019; Zimmer *et al.* 2019). The histogram of the oriented phase congruency (HOPC) based on shape features is successfully used to match the multisource remote sensing images (Ye *et al.* 2017). Generic and automatic Markov Random Field is defined by grayscale and gradient statistical information (Yan *et al.* 2018). The main ideal of robust center-symmetric local ternary pattern (CSLTP) based self-similarity descriptor is a rotation invariance description strategy on local correlation surface (Chen *et al.* 2018). Extended phase correlation algorithm based on log Gabor filtering (LGEPC) focuses on two problems: 1) significant nonlinear radiometric differences and 2) large-scale differences between image pairs. (Xie *et al.* 2019). These methods are robust to the significant nonlinear intensity differences between multisource remote sensing images. However, they are may limited to the terrain structure information.

In an appropriate region, the terrain shapes are extremely similar between visible and SAR images (called scene shape similarities) regardless of whether they have highly different intensity characteristics. Inspired by this discovery, this study proposes a multispectral image matching strategy based on scene shape similarity. Therefore, shape context can be applied to describe scene contour structures (Belongie *et al.* 2001). Accordingly, a novel feature descriptor is defined by the shape features that can describe the scene shape structures of the images. The descriptor, namely the scene shape similarity feature (SSSF), can be efficiently calculated using the shape context algorithm over the image. The SSSF descriptor reflects the terrain shape structural properties of a certain region in images, which are robust to the intensity pattern between two images. The SSSF descriptor can be calculated for the reference and the sensed image. The NCC of the SSSF descriptor is then used to define the similarity measure, called the $SSSF_{ncc}$, which is used to extract tie points via a template-matching method. Finally, the PL transform is used to rectify the sensed image. The main contribution of this study is the definition of a novel scene shape similarity measure named $SSSF_{ncc}$, which can handle nonlinear grayscale differences between multisource images.

The proposed method is introduced in the following parts: Section "Methodology" describes the methodology, including

shape context, the SSSF descriptor, similarity measures based on the SSSF descriptor, and multisource registration methods based on $SSSF_{ncc}$. Then, Section "Experiments and Evaluation" presents the experiments and evaluates the results. Section "Discussion" summarizes the conclusions and discussions.

Methodology

Image registration aims to capture the optimal alignment under the spatial coordinate system between multisource remote sensing images. This study proposes a robust multispectral remote sensing image registration method based on scene shape similarity. In this method, the blocked Harris operator is first applied to detect feature points in the reference image. Then, the novel SSSF descriptor is designed using the shape context algorithm. Furthermore, a similarity measure defined by the SSSF descriptor is applied to extract tie points via the template matching method. A global consistency check method is then used to remove the mismatched tie points. Finally, the PL transform model is selected to rectify the remote sensing images. Figure 1 shows the key procedure of the proposed method.
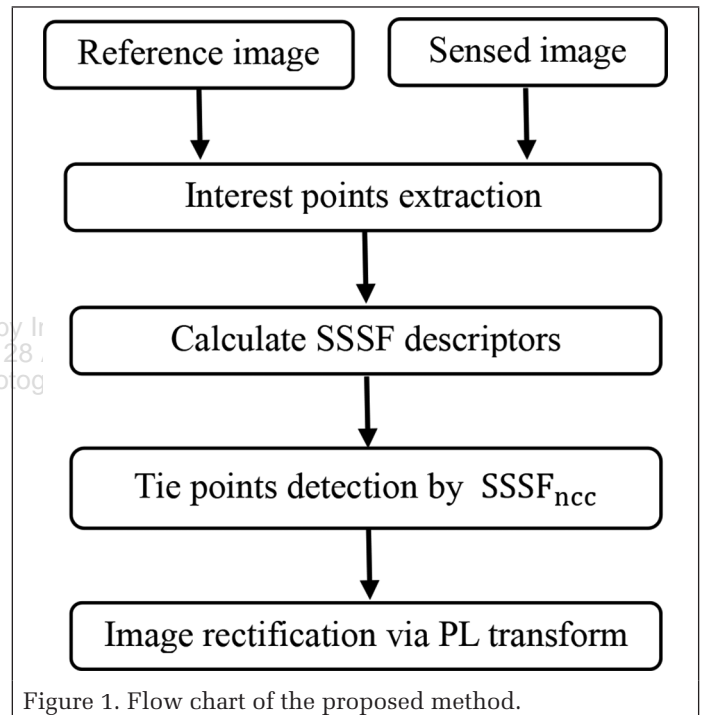

Figure 1. Flow chart of the proposed method.

Shape Context

Shape context is a very popular shape descriptor and is mostly used for shape matching and target recognition. It adopts a feature description method based on shape contours, which uses the histogram to describe the shape features in the log polar coordinate system to reflect the contour's distribution of sampling points. In the overall framework, the shape context counts the context information of each point in a two-point set and compares the context information between each point to obtain the closest set of permutations. Moreover, the correspondence points of the second point set can be found for each point of the first point set. The basic principle of the shape context is as follows:

1. For a given shape, the shape contour is obtained through the edge detection operator (e.g., canny operator). Then, the contour of the given shape is sampled to obtain a set of discrete points $p_1, p_2, ..., p_n$. Figures 2a and 2b show this detail.

2. The shape context is calculated. Any point p_i is taken as a reference point. N concentric circles are established at a logarithmic distance interval in a local area, where p_i is the center. This region is equally divided in the circumferential direction M to form a target template, as shown in Figure 2c. The vector relative position of point p_i to other points is reduced to the number of point distributions in each sector on the template. The statistical distribution histogram $h_i(k)$ of these points, called the shape context of point p_i, is calculated as:

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in \text{bin}(k)\} \quad (1)$$

where $k = \{1, 2, …, K\}$, $K = M \times N$. In this paper, the concentric circle is set to five layers, and the circle is divided into 12 partitions (N = 5, M = 12).

The use of logarithmic distance segmentation allows the shape context descriptor to enhance local features and be more sensitive to adjacent sample points rather than to away from the point. As shown in Figures 2d and 2e, the shape context at different points of the contour varies, but the corresponding points of similar contours tend to have similar shape contexts.

For the entire point set P, the n points $p_1, p_2, …, p_n$ are used as reference points, and the shape histograms composed of the remaining n − 1 points are sequentially calculated. Finally, n-shaped histograms are obtained and stored in a matrix of size n×(n−1). Thus, for any target, a matrix of size n×(n−1) can be used to represent its shape information, and a matrix of size n×(n−1) is the shape context of the set of points P, therefore describing the characteristics of the entire contour shape. Finally, the shape context of the given shape is obtained through Steps 1 and 2.

In the template window extended by feature points, a concentric circle with each point as the center point is drawn. And the concentric circle is set to five layers, the circle is divided into 12 partitions, so there are 5×12=60 grids. Then the distribution of other points in these 60 grids should be counted. This distribution is presented by sampling the angles and distances of other points relative to the feature points in a polar coordinate system. Finally, an array of length 60 is needed to record the distribution of other points around the feature points. This completes the sampling of the contours in the template scene.

SSSF Descriptor
This thesis mainly aims to define a robust descriptor that can function despite the significant nonlinear grayscale differences between multisource remote sensing images. In this section, the shape context algorithm is used to propose a novel

descriptor named SSSF. The SSSF descriptor characterizes the terrain structure similarity of images using shape context. The shape context feature is reliable for grayscale differences and intensity changes and also reflects the local shape feature of images. Therefore, the shape context feature is utilized to build the SSSF descriptor. Figure 3 shows the process of extracting the SSSF descriptor, and the details of each step are as follows:

1. The first step extracts a feature point from the feature point set.

2. The second step extends a certain size template window that centers on the feature point and extracts template window boundaries using the canny operator. This step is for calculating the SSSF descriptor.

3. The third step calculates the histogram of shape context information over the template window. Here, only the shape context at the center point of the template window (i.e., feature point) is calculated to reduce computational time. Each histogram for the template window is normalized by the L2 norm to adjust against illumination changes well. This process calculates the shape context for the template window and obtains the scene shape feature information.

4. The final step builds the SSSF descriptor by reshaping the histogram of the template window. This descriptor will be used for multisource image registration.

The SSSF descriptor inherits the characteristics of the shape context algorithm. In the target region centered on the feature
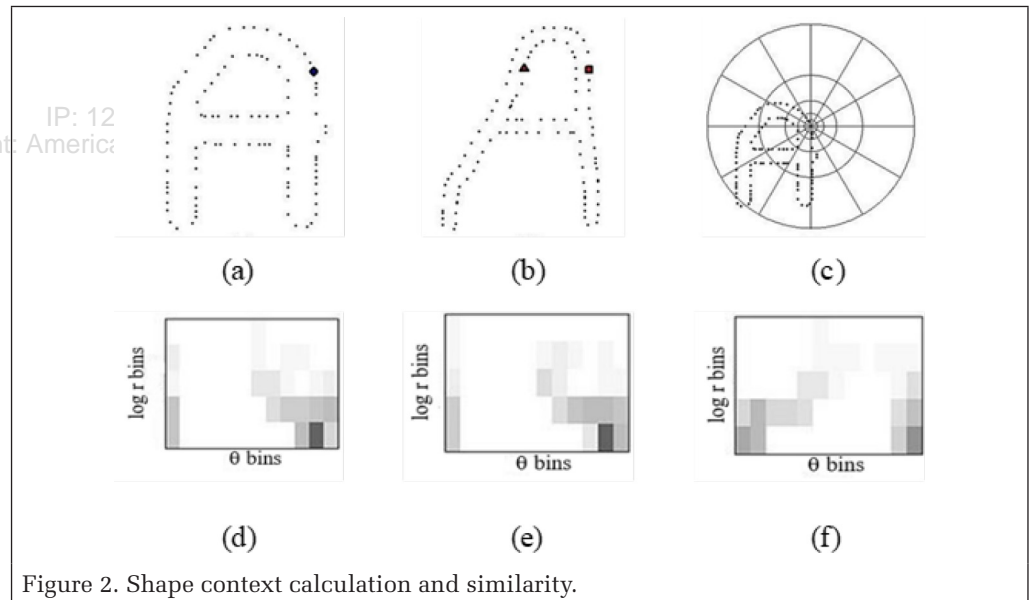

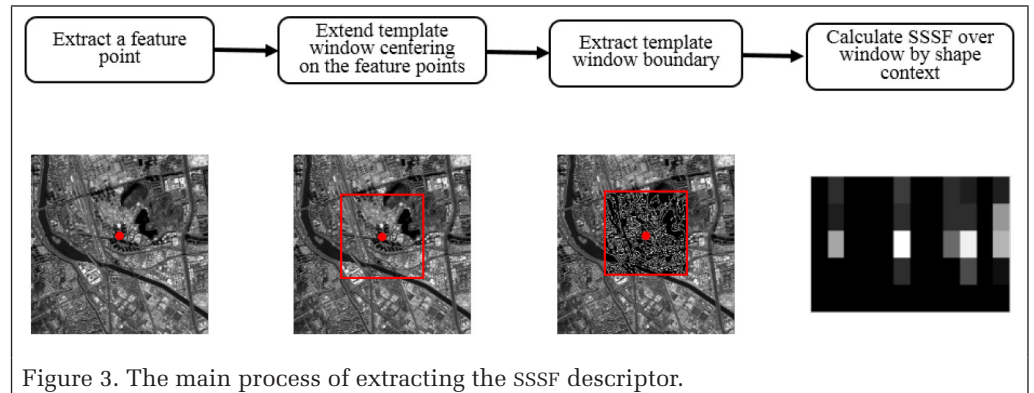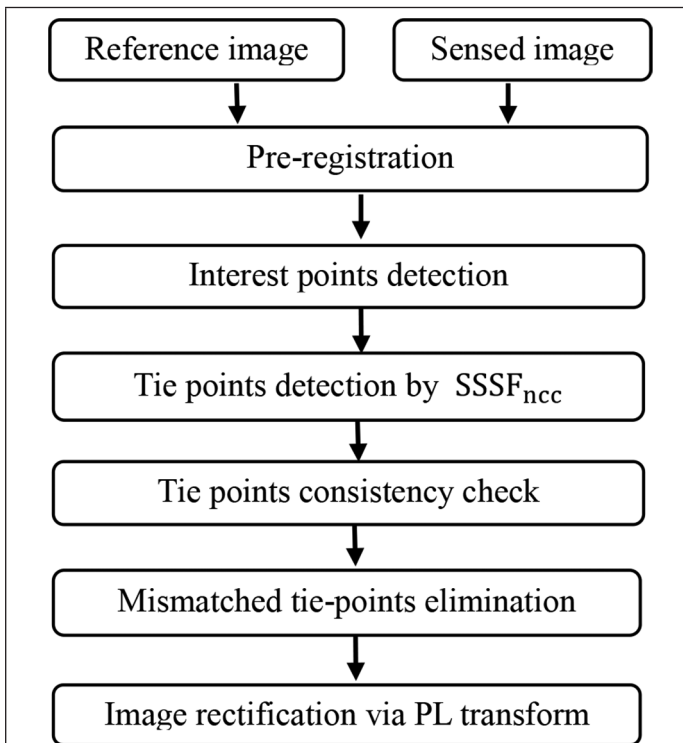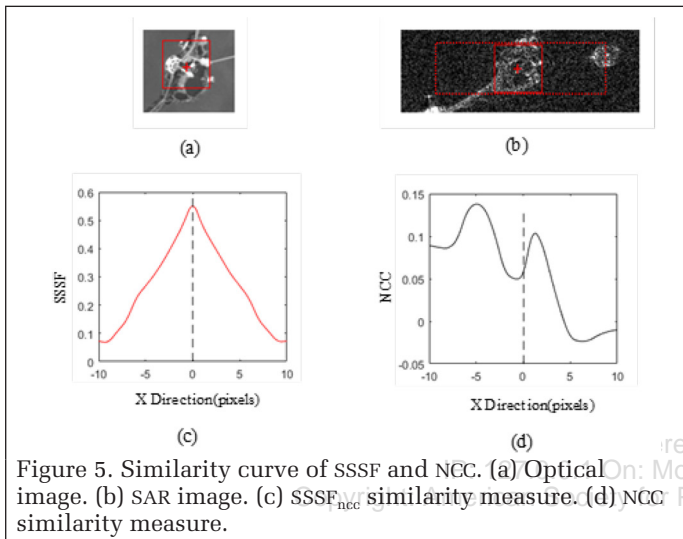Figure 2. Shape context calculation and similarity.

Figure 3. The main process of extracting the SSSF descriptor.

Figure 4. SSSF descriptor computed from the tie point of the optical and SAR images in the same region.

Figure 5. Similarity curve of SSSF and NCC. (a) Optical image. (b) SAR image. (c) SSSF_ncc similarity measure. (d) NCC similarity measure.

Figure 6. Main matching procedure of the proposed method.
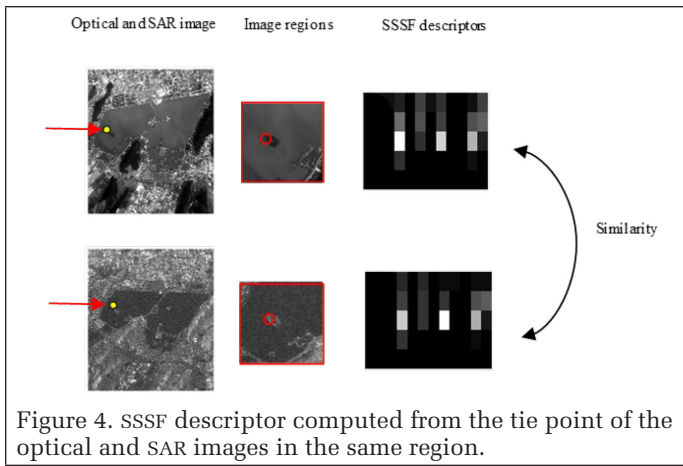
point, the number of contour points falling within each sector is counted to build the histogram of shape context information under polar coordinates. The SSSF descriptor is built by sampling the histogram in the polar coordinate system to the Cartesian coordinate system. The horizontal and vertical coordinates are expressed in a radial (N = 5) and angular direction (M = 12), respectively. This descriptor reflects the distribution of other points around the feature point. In other words, it reflects the characteristic of the terrain structure in the template window. Therefore, the similarity between SSSF descriptors is used as a metric to match multisource images.

Similarity Measure Based on the SSSF

As mentioned above, the SSSF descriptor can detect the scene shape structures of a certain region in images, making it reliable for significant nonlinear grayscale differences. Figure 4 shows that the SSSF descriptor is computed from the tie point of the optical and SAR images in the same region. Although significant grayscale differences exist between the two images, the descriptors are quite similar. On the basis of the similar shape feature, this descriptor can be applied to multisource remote sensing image registration. Therefore, this study uses the NCC of the SSSF descriptor (abbreviated as SSSF_ncc) as the similarity measure for multisource image registration. The SSSF_ncc is calculated as follows:

$$\text{SSSF}_{\text{ncc}} = \frac{\sum\limits_{k=1}^{n}(h_p(k)-\overline{h_p(k)})(h_q(k)-\overline{h_q(k)})}{\sqrt{\sum\limits_{k=1}^{n}(h_p(k)-\overline{h_p(k)})^2\sum\limits_{k=1}^{n}(h_q(k)-\overline{h_q(k)})^2}}, \quad (2)$$

where p and q are the point features of the two images, respectively; $h_p(k)$ and $h_q(k)$ are the SSSF descriptors of a template region centered at p and q, respectively; $\overline{h_p(k)}$ and $\overline{h_q(k)}$ indicate the means of the SSSF descriptors.

SSSF is compared with NCC through the similarity curve to reveal its advantages in terms of matching multisource images. The test selects a set of optical and SAR images with significant nonlinear grayscale differences. First, a template window centering on the feature points with a certain size (typically 15×15 pixels) is extended. Then, the NCC and SSSF are computed within a search region (20×20 pixels) of the SAR image. The similarity curve of SSSF and NCC is shown in Figure 5. Comparison shows that SSSF obtains the correct match point, whereas NCC fails to obtain tie points. This example indicates that SSSF is preliminarily more robust to significant intensity differences compared with NCC.

Multisource Image Registration based on SSSF_ncc

This section introduces a reliable multispectral image registration method based on SSSF_ncc. Figure 6 shows the main matching procedure of the proposed method. The detailed steps are as follows:

1. The obvious rotation differences between two images are first removed via coarse rectification. Then, the possible resolution differences can be eliminated by resampling the two images into the same ground sample distance (GSD). In addition, the SAR images used in this experiment have not been subject to pretreatment, such as radiometric calibration and topographic correction. However, for SAR images with different azimuth and range resolutions, multi-view processing is required to eliminate the deformation of the ground caused by SAR side-view imaging.

2. The blocked Harris operator is used to obtain uniformly distributed feature points from the reference image. Then,

the reference image is divided into some regular grids without overlap. The Harris intensity value H of each pixel in each grid is calculated. The H values are arranged in descending order, and the first n points are selected as feature points.

3. After the feature point is detected in the reference image, a corresponding search area (typically 20×20 pixels) of the sensed image is predicted by the known geographic coordinate information of the images. Then, the SSSF$_{ncc}$ is applied to extract the tie points via a template matching method in this search region. Although the semiglobal matching algorithm has been commonly used in a dense match, this method may be not suitable for the method of this study and is slightly time consuming (Hirschmuller *et al.* 2007). A bidirectional matching technique, which consists of two steps (positive and reverse matching steps), is applied to image matching to acquire a robust match result (Qu *et al.* 2016). In the positive matching step, for a point p extracted from the feature point sets of the reference image, the corresponding feature point q can always be found through the maximum similarity of SSSF$_{ncc}$ between the template window and the search region. In the reverse matching step, the corresponding point of q in the reference image is obtained through the same strategy. When the two steps receive an identical match result, the matched point pair (p, q) is regarded as tie points.

4. Several uncertainty factors should be considered in the image matching process, for example, the occlusion and shadow. Therefore, the above-extracted tie points are not always accurate. The projective transformation model is usually used for the tie point consistency check, because this model can deal with conventional global transformation (such as rotation, scale, and translation) effectively.

5. The steps of tie points consistency check are as follows: An iterative refining procedure is used to eliminate mismatched tie points. A projective transformation model is first calculated using the least-squares method with all the tie points. Then the residual errors and the root-mean-square error (RMSE) of tie points are computed to evaluate matching accuracy. The tie point with the largest residual error is iteratively removed. The aforementioned process is repeated until RMSE is less than a given threshold (e.g. 1 pixel).

6. After the mismatched tie points are removed, the triangulation-based PL model is applied to fit the local distortions caused by terrain relief. For the images covering mountain and hilly regions, the first, second order polynomial transformation models are better than PL transform model for prefitting the nonrigid deformations among such images. However, the test data is located in the plains, without obvious undulating terrain. Therefore, in this paper, the PL model is used to rectify the sensed images. This model first constructs the triangulated irregular networks (Delaunay TINs) (Ye *et al.* 2015) with all the tie points. Then, the transformation parameters in each triangular region are calculated by the coordinates of the triangle's three vertices. Finally, the image is rectified on each triangle region.

$$x_p = a_0 + a_1 x_q + a_2 y_q$$
$$y_p = b_0 + b_1 x_q + b_2 y_q$$

where (x_p, y_p) and (x_q, y_q) are the coordinates of the reference and remote sensing images, respectively.

Experiments and Evaluation

Experiments are conducted on five pairs of multisource remote sensing images from optical, SAR, and map data to evaluate the proposed method. The test sets consist of three parts: high-resolution and medium-resolution images and image-to-map sets. Each set consists of an image pair composed of a reference and a remote sensing image, which are obtained from different sensors with significant nonlinear grayscale differences. The test data, implementation details, and experimental analysis are discussed in the following pages of this paper.

Description of Test Sets

Three categories of multisource remote sensing image pairs are used to evaluate the effectiveness of SSSF$_{ncc}$. Table 1 presents the descriptions of test sets, and Figure 7 shows the test data images. The details of each test set are as follows:

High-Resolution Set
In the reference data (optical image), the multispectral red-light band data of China Resources No. 3 satellite image is used. The spatial resolution is 6 m below the star point, and the image acquisition time is dated June 2017. The remote sensing data (SAR image) is based on the Japanese Advanced Land Observing Satellite (ALOS) Earth observation satellite Phased Array L-Band Synthetic Aperture Radar (PALSAR) sensor spatial resolution. The rate is 6 m, and the image time is dated June 2012. The coverage area of the optical-SAR 1 is mainly near the Yunlong Lake in Xuzhou, China, which is also mixed with some road, house, and other land information. The area is located in the plains, without obvious undulating terrain, and the local distortion is significant at the riverbank and high buildings. The coverage area of the optical-SAR 2 is Xuzhou's Yunlong District, with various houses, buildings, and roads with richly shaped features. Similar to the Yunlong Lake, no obvious undulating terrain is found in the city. However, high buildings clearly lead to local deformation. These two data sets include significant local deformations affected by building displacement due to the capture from different sensors. In addition, the two images have a five-year time difference, and a few ground objects have changed during this period. The rotation difference can be observed between two images (see Figures 7a and 7b).

Medium-Resolution Set
The reference data (optical image) come from the Band3 red band data on the Landsat5 TM sensor of the U.S. Landsat series. The spatial resolution is 30 m and the image acquisition time is dated January 2003. Data (SAR image) come from the European Space Agency Earth observation satellite series ENVIronment SATellite (ENVISAT) satellite Advanced Synthetic Aperture Radar sensor. The image acquisition time is dated March 2004, and the resolution is 20 m. The experimental area of the optical-SAR 3 is near the Yunlong Lake in Xuzhou. The test area of the optical-SAR 4 is the Xuzhou Yunlong District. The images have a temporal difference of almost 14 months. Figures 7c and 7d show that the terrain shape structures of the images have lower quality than the high-resolution set, and the local deformation caused by high buildings and terrain changes is not as significant.

Map Set
The reference image is obtained from the Baidu map, whereas the remote sensing image is obtained from Google Earth. These images were acquired in December 2016, and their resolution is 2 m. As shown in Figure 7e, the map data have less terrain shape structure information, and the radiometric properties and local distortions are evident between the two images. Therefore, detecting the tie points between the two data is challenging.

Table 1. Descriptions of test data.

Category	Test	Reference Image			Sensed Image		
		Sensor	Size and GSD	Date	Sensor	Size and GSD	Date
High-resolution set	Optical-SAR 1	ZY-3	1000 × 900, 6 m	2017.06	ALOS	1000 × 900, 6 m	2012.06
	Optical-SAR 2	ZY-3	1000 × 900, 6 m	2017.06	ALOS	1000 × 900, 6 m	2012.06
Medium-resolution set	Optical-SAR 3	Landsat5	594 × 594, 30 m	2003.01	ENVISAT	594 × 594, 30 m	2004.03
	Optical-SAR 4	Landsat5	594 × 594, 30 m	2003.01	ENVISAT	594 × 594, 30 m	2004.03
Map set	Image-map	BAIDU	695 × 457, 2 m	2016.12	Google Earth	695 × 457, 2 m	2016.12

Figure 7. Images for all tests: (a) optical-SAR 1, (b) optical-SAR 2, (c) optical-SAR 3, (d) optical-SAR 4, (e) map set.
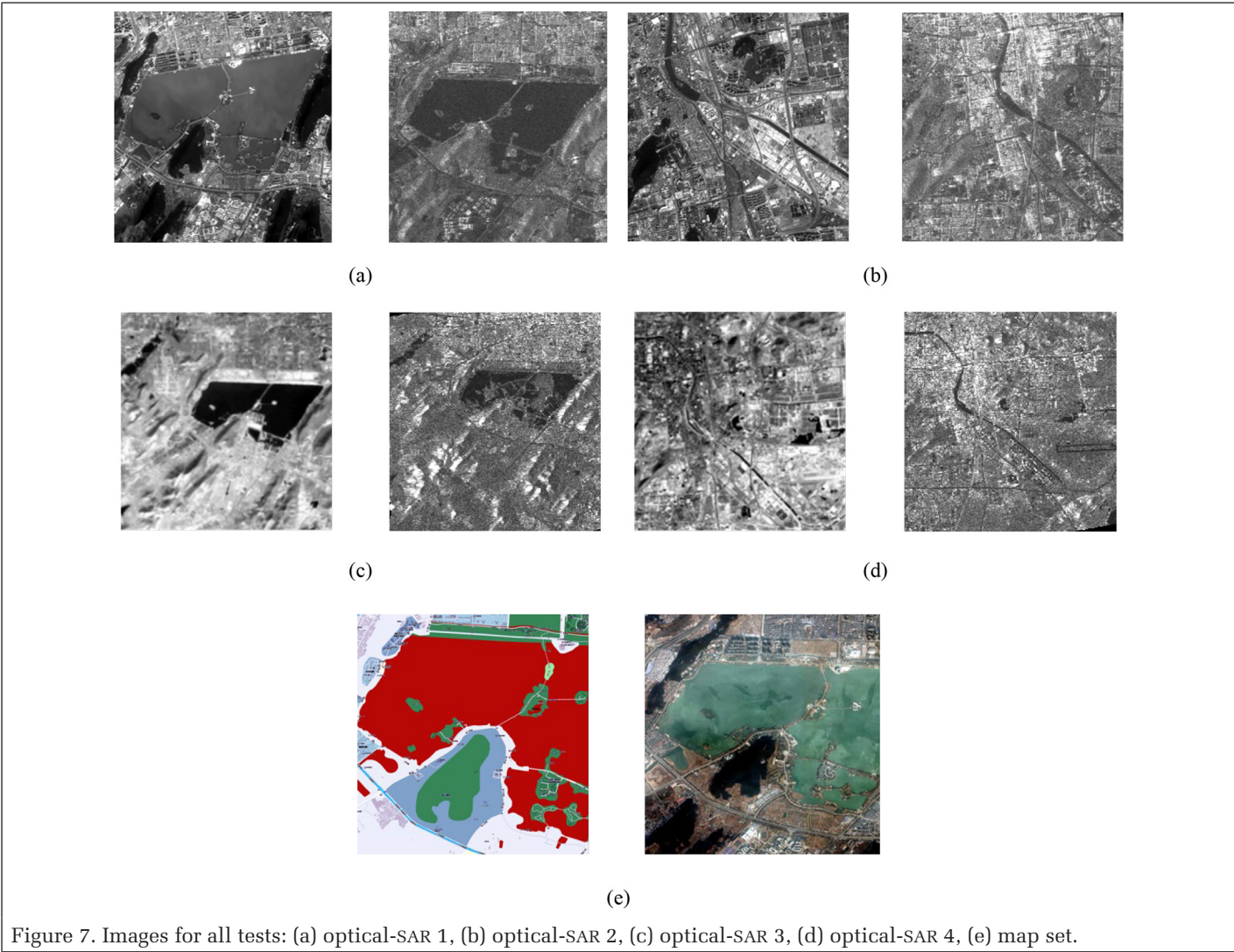
Parameter Settings

For all experimental tests, the blocked Harris operator was applied to extract the feature points from the reference image. The reference image was first divided into 10×10 blocks, and eight Harris feature points were extracted in each block. This setup detected a total of 800 feature points. The SSSF descriptor is computed by setting the template window to different sizes (i.e., from 5×5 pixels to 40×40 pixels) to analyze the sensitivities of $SSSF_{ncc}$ with regard to the changes of template window size. During this process, different canny operator thresholds (i.e., from 0.1 to 0.3) were set to obtain different sharpness levels in the contours of the template window. For the consistency check method, the projective model was applied to check the mismatched points of the high-resolution set because of the few local distortions. By contrast, for the medium-resolution set and map set, the third-order polynomial was selected for error eliminations due to several local distortions. Finally, the RMSE threshold of all the tests was set to 1 pixel to evaluate the matching accuracy.

According to the data listed above, the construction process of the SSSF descriptor has two key parameters: the template window size and the canny operator threshold. Therefore, different parameters are set to analyze the sensitivities of $SSSF_{ncc}$ concerning changes of the template window size and canny operator threshold.

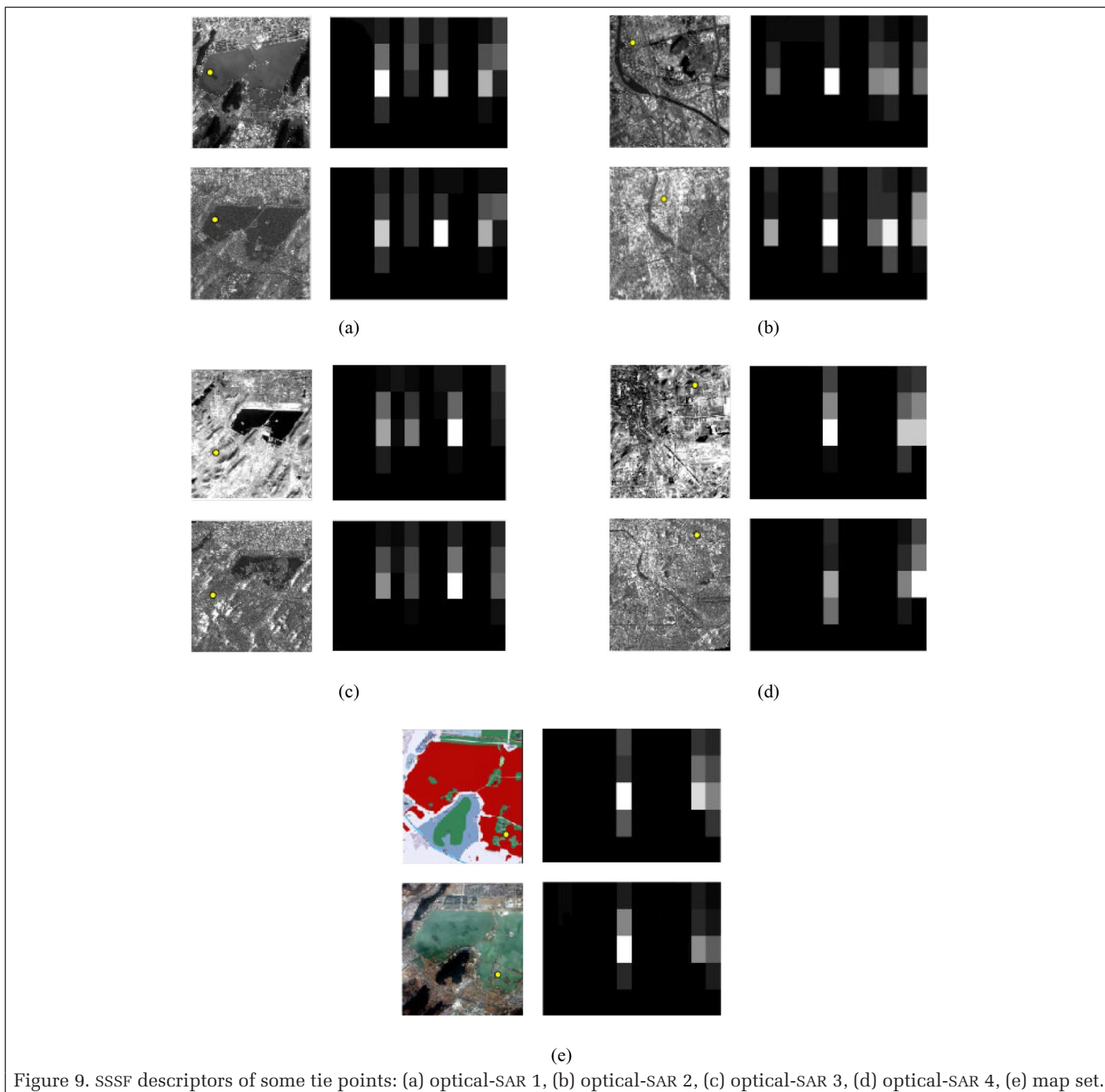## Experimental Details and Assessment Criteria

In the experiments, the blocked Harris operator was first applied to detect the feature points in the reference image. A total of 800 feature points were evenly detected over the reference image. Then, one feature point was extracted from the

Figure 8. Example of terrain contours from optical-SAR 2.
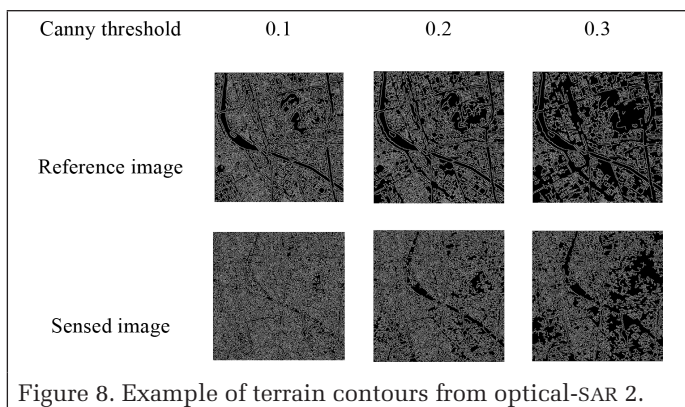
point set. A template window centering on the feature points with different sizes (from 5×5 pixels to 40×40 pixels) was extended to analyze the sensitivities of SSSF_{ncc} with regard to changes of the template window size. At the same time, different canny operator thresholds (from 0.1 to 0.3) were set to obtain different sharpness contours of the template window (see Figure 8).

In accordance with the steps listed above, the SSSF descriptor of the central point can be calculated using the shape context algorithm (see Figure 9). In the matching procedure, a search window of 20×20 pixels in the image was first predicted by the known geographic coordinates, and the SSSF_{ncc} was applied to extract the tie points via the template matching strategy.

Figure 9. SSSF descriptors of some tie points: (a) optical-SAR 1, (b) optical-SAR 2, (c) optical-SAR 3, (d) optical-SAR 4, (e) map set.
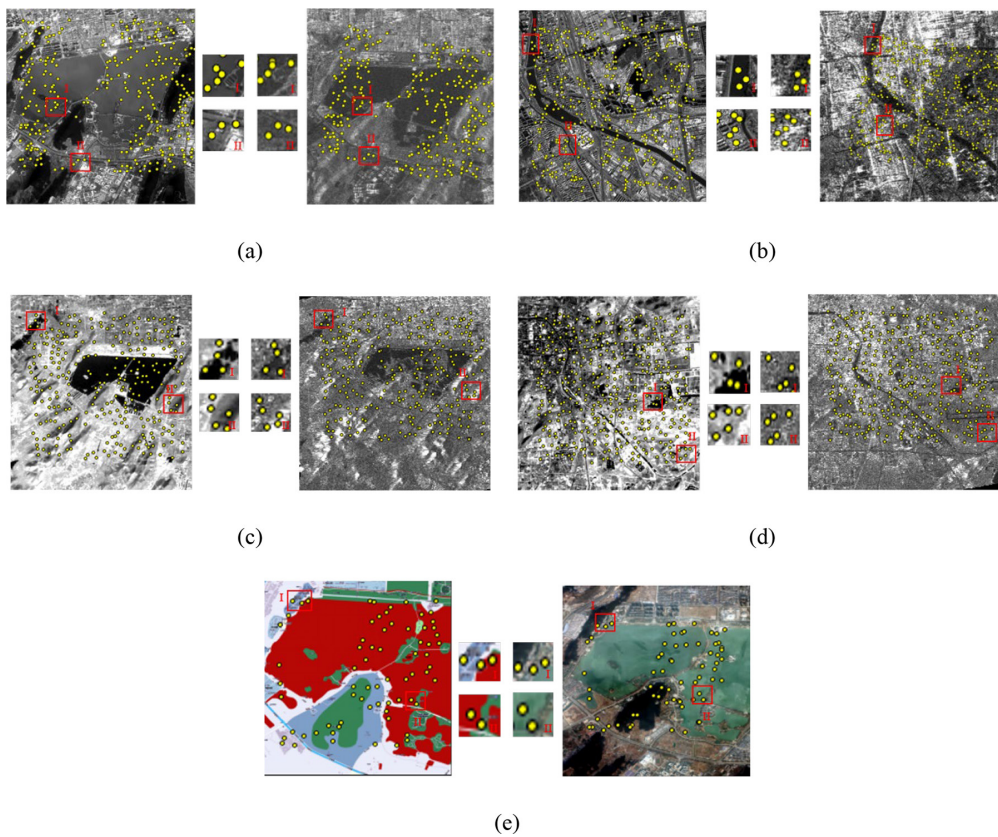
Figure 10. Tie points detection for all tests: (a) optical-SAR 1, (b) optical-SAR 2, (c) optical-SAR 3, (d) optical-SAR 4, (e) map set.
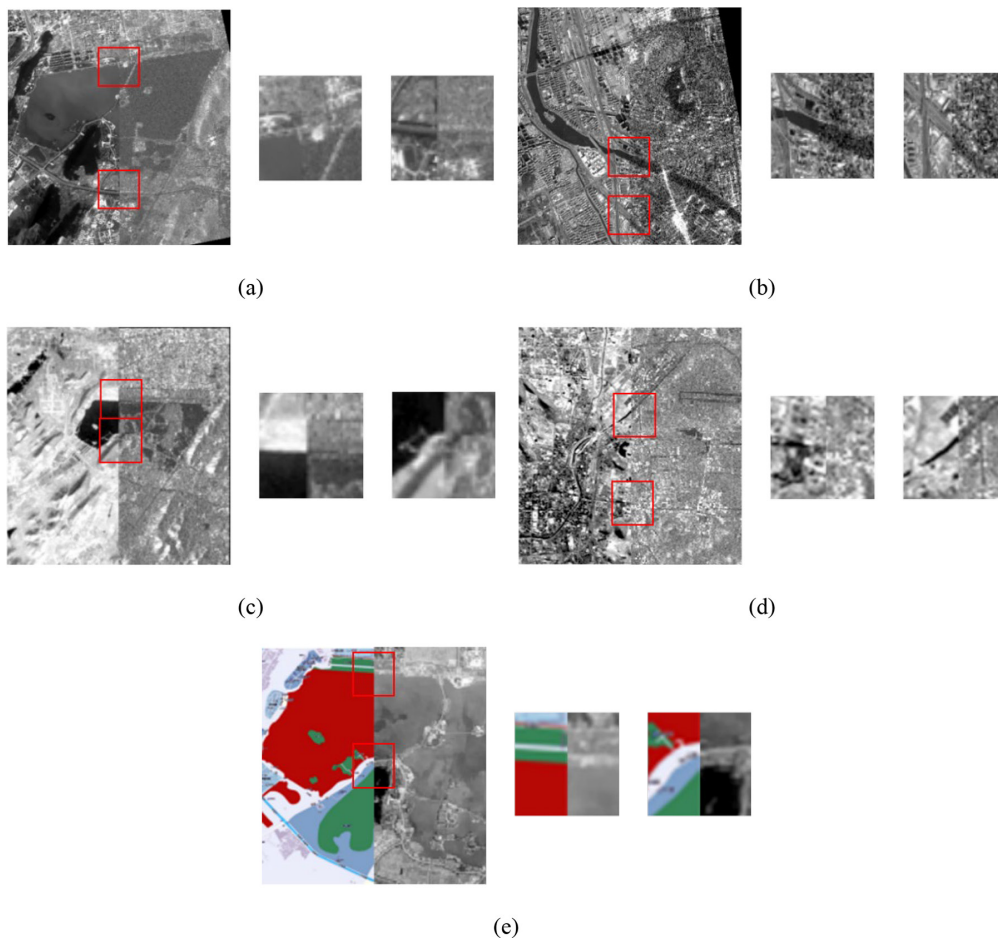

Figure 11. Registration results of all the test sets: (a) optical-SAR 1, (b) optical-SAR 2, (c) optical-SAR 3, (d) optical-SAR 4, (e) map set.

Correct match rate (CMR) is calculated as CMR = CM/C, where CM is the number of correctly matched point pairs, and C is the total number of match point pairs. For each image pair, 30 evenly distributed check points are first manually selected across the two images. Then, the residual error of the matched point pairs is computed using a geometric transformation model. The point pair is considered the correct match to determine the CM, only when the residual error of the point pair is less than 1.3 pixels. Moreover, the RMSE (threshold is 1 pixel) of the correct matched tie points is used to assess the registration accuracy. After the mismatched points are removed, the tie points between the multisource remote sensing images are extracted (see Figure 10). Finally, the multisource registration is achieved through the PL transformation model (see Figure 11).

Experimental Analysis

Accuracy Analysis
The match accuracy of the proposed method is assessed by two methods. One method is a visual judgment to verify the splice of the matched and reference images. Another method is to calculate the RMSEs of check points between the two images.

The registration results of the five test sets are shown in Figure 11, and the subimages contain different terrain structures, including lake banks, roads, and rivers. As shown in the subimages, the matched image coincides with the reference image well regardless of the nonlinear grayscale differences. The slight deviations at the road splices (the second subimage in Figure 11b) are still visible because of the road displacement. The high-resolution image match strategy can be handled fundamentally by an image-to-image matching method only when a true orthostatic rectification is applied.

Meanwhile, a set of 30 manually selected check points is used to compute the RMSEs of the five tests. The performance of image registration, as indicated by the RMSEs, is shown in Figure 12 and Table 2.

As presented in the section "Parameter Settings", different parameters are set to analyze the sensitivities of $SSSF_{ncc}$ concerning the changes of the template window size and canny threshold. Figure 12 shows the RMSEs of each test of correct matched point pairs. As the size of the template window increases, the RMSEs of these tests initially rises then falls. The image matching accuracy of each test is the highest when the template window size is 15×15 pixels. Moreover, when the canny threshold is set to 0.2, the proposed method obtains an improved matching outcome. The reasons for these results are as follows:

For the template window size, any terrain feature in a small region cannot be extracted. Additionally, the time and geometric differences, local distortion, and the increase/decrease and deformation of terrains across a large region vary considerably, thus resulting in the drop of scene shape similarity. Comparison shows that the terrain shape structures of appropriate regions are quite similar.
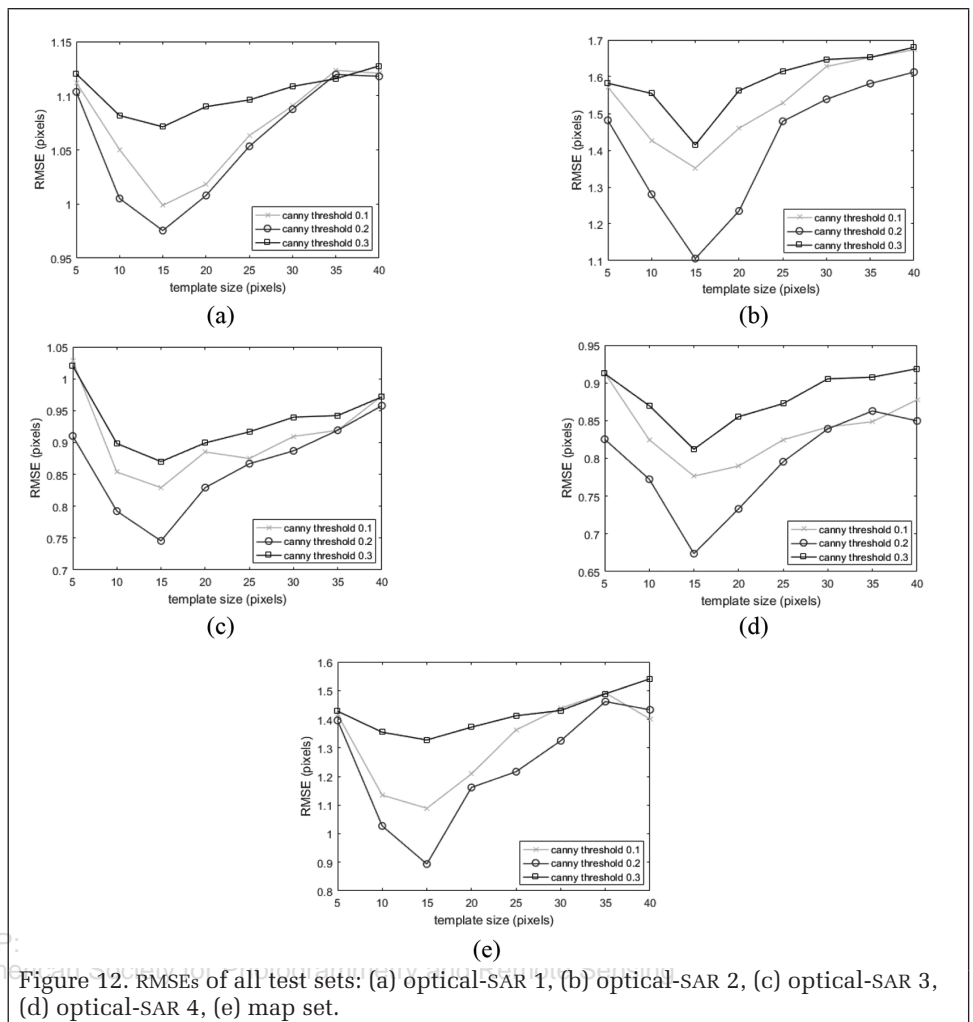

Figure 12. RMSEs of all test sets: (a) optical-SAR 1, (b) optical-SAR 2, (c) optical-SAR 3, (d) optical-SAR 4, (e) map set.

Table 2. Matching performance of all methods.

Test	Methods	Matched Points	CMR (%)	RMSE (Pixel)	Time (Seconds)
Optical-SAR 1	Proposed	253/286	88	0.97	93.1
	NCC	79/233	33	1.5	28.3
	HOPC$_{ncc}$	158/184	86	0.976	107.9
	CSLTP	208/247	84	0.99	88.4
Optical-SAR 2	Proposed	442/491	90	1.11	87.5
	NCC	161/436	37	1.6	24.1
	HOPC$_{ncc}$	186/214	87	1.192	94.9
	CSLTP	381/448	85	1.3	81.7
Optical-SAR 3	Proposed	222/244	91	0.74	35.6
	NCC	67/193	35	1.4	16.7
	HOPC$_{ncc}$	127/141	90	0.809	39.4
	CSLTP	195/224	87	0.93	33.5
Optical-SAR 4	Proposed	327/352	93	0.67	36.1
	NCC	114/301	38	1.3	14.6
	HOPC$_{ncc}$	163/171	95	0.806	36.9
	CSLTP	291/323	89	0.72	34.8
Image-map	Proposed	77/85	90	0.89	25.1
	NCC	18/63	28	1.4	12.2
	HOPC$_{ncc}$	65/82	79	0.912	28.9
	CSLTP	62/76	81	0.97	23.3

Matched points: correct match points/total match points.
CMR: correct match rate.

Figure 8 shows the optical-SAR 2 contours detected by a different canny threshold. The image contours detected by the canny threshold of 0.1 are too dense and plentiful to reduce the impact of the increase/decrease and deformation of terrain. By contrast, contours are not enough to use the canny threshold of 0.3 to extract the terrain shape features, making it difficult to extract a sufficient number of tie points. Therefore, the experiments achieve the highest registration accuracies when the parameters are set to 15×15 pixel template windows with a canny threshold of 0.2. The remaining experiments are implemented on the basis of these parameters (include the results shown in Figure 9, Figure 10, Figure 11, and Table 2).

Table 2 shows that the proposed SSSF$_{ncc}$ has been successfully applied to match all data. These tests achieve different registration accuracies due to terrain and resolution differences. The registration accuracies of the medium-resolution and map sets are higher than the high-resolution set. The registration results have such significant differences that the medium-resolution set with low resolution contains small local distortions. The map set data are obtained from simplified real terrain (classify terrains into one class) containing small local distortions. High-resolution sets obtain the lowest registration accuracy due to the increase/decrease and clear deformation of terrain between two images. However, the RMSEs of optical-SAR 1 and optical-SAR 2 are 0.97 and 1.11 pixels, respectively. These values are acceptable accuracies for the registration of high-resolution images.

The number of tie points and the CMR have some differences among the five tests. The optical-SAR 2 and 4, which contain urban areas, have more terrain features compared with optical-SAR 1 and 3, which contain a large water area. Therefore, more tie points can be detected in the optical-SAR 2 and 4 images. As for the map set, detecting too many tie points is difficult due to the lack of terrain. The CMRs of all five groups of data are very high. Compared with optical-SAR 1 and 3, the CMRs of optical-SAR 2 and 4 are higher due to the similarity and amount of terrain between the reference and sensed remote sensing images. Therefore, the optical-SAR 4 achieves the best registration accuracy. In summary, these registration results show that the SSSF$_{ncc}$ similarity measure is robust for significant nonlinear grayscale differences among multisource remote sensing images.

Comparison of SSSF with Other Methods
In this study, the SSSF$_{ncc}$ similarity measure is used to detect tie points. The proposed SSSF$_{ncc}$ is compared with the NCC, HOPC$_{ncc}$, and CSLTP to illustrate its accuracy in terms of matching multisource remote sensing images.

The histogram of the oriented phase congruency (HOPC) (Ye *et al.* 2017) based on shape features is successfully used to match the multisource remote sensing images. This descriptor first divides the image into several blocks, with each block comprising several "cell" units. The phase consistency histograms in all cells are counted and linked together to form the final feature description vector. The basic idea of phase consistency is that the image features always appear at the maximum phase overlap of the Fourier harmonic components. Then, the correlation coefficient between the feature vectors (HOPC$_{ncc}$) is used to match the correspondence points. However, this descriptor is limited by the structural features of the image and can be time consuming.

The main ideal of CSLTP based on self-similarity is a rotation invariance description strategy on local correlation surface. The SIFT algorithm is applied for local feature detection and the CSLTP descriptor is constructed for each extracted feature points. Then, a bilateral matching strategy combined with an outlier removal procedure in the geometric transformation model are applied for feature registration and mismatch elimination. This method has successfully matched various multisource satellite images and results show that it is inherently rotation invariant and robust to complex intensity differences. This study compares the HOPC$_{ncc}$ and CSLTP to illustrate the reliability of the proposed method.

Five comparison experiments were set to verify the accuracy of SSSF$_{ncc}$ by detecting tie points between multisource images. First, the same 800 Harris feature points were detected in the reference image. The tie points were extracted in a search region (20×20 pixels) of the sensed image via a template matching method using the SSSF$_{ncc}$, NCC, HOPC$_{ncc}$ as measures, while the CSLTP uses a bilateral matching strategy to match tie points. The most appropriate parameters were set to three similarity measures to analyze the registration accuracy of these three similar measures. For example, the HOPC$_{ncc}$ was set with the parameters of $\beta = 8$ orientation bins, 3×3 cell blocks of 4×4 pixels cells, and a $\alpha = 1/2$ block width overlap. A global consistency check method was then used to remove the mismatched tie points.

Table 2 lists the matching performance of all the five tests, including the total matches, matched points, CMRs, matching times, and RMSEs. The times, CMRs, and RMSEs are applied to assess the registration accuracy of the proposed method. Figure 13 shows the times, RMSEs, and CMRs of these three similar measures.

As shown in Figure 13 and Table 2, the RMSEs and CMRs of SSSF$_{ncc}$ show better results compared with the other methods for all five tests. The analysis is as follows: The NCC is robust for linear intensity patterns rather than the complex nonlinear grayscale differences across multisource images. The HOPC$_{ncc}$ also achieves good registration outcome. However, the CMRs and RMSEs of HOPC$_{ncc}$ are slightly lower than the proposed method, because the HOPC$_{ncc}$ is defined by the single amplitude and orientation of phase congruency (Ye *et al.* 2017), whereas the SSSF$_{ncc}$ is based on the relative position of the vector from the center of the area to the other points. The accuracy of CSLTP is lower than SSSF$_{ncc}$ and HOPC$_{ncc}$ may be because it is inherently rotation invariant rather than robust to the local deformation caused by high buildings and river
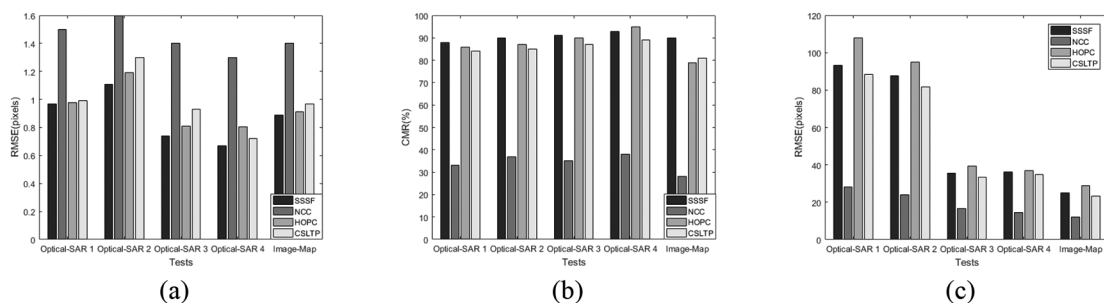

Figure 13. (a) RMSEs, (b) CMRs, and (c) Times of three similarity measures.

bank. Therefore, the SSSF_{ncc} achieves more comprehensive and complete terrain shape features compared with HOPC_{ncc} and CSLTP in a region. Besides, why the HOPC_{ncc} gets the minimal tie points is that the HOPC_{ncc} removes image boundaries to eliminate the effects of noise and illumination.

For computational efficiency, the NCC has the fastest calculation. SSSF_{ncc} computes faster than HOPC_{ncc}, because SSSF_{ncc} and HOPC_{ncc} have to integrate the descriptors first and then compute the NCC between these descriptors, which is time consuming. SSSF_{ncc} requires less run time compared with the HOPC_{ncc}, mainly because the SSSF_{ncc} only calculates the shape context centering on the feature points. This process is less time consuming than computing the phase congruency feature applied to build the HOPC_{ncc}. The results in Figure 13 illustrate that SSSF_{ncc} is more time consuming than CSLTP mainly because SSSF_{ncc} is required to extract scene shape information and calculate the NCC between descriptors, which needs more run time than calculating self-similarity used to construct CSLTP. Overall, results illustrate that compared with the NCC, HOPC_{ncc} and CSLTP, the SSSF_{ncc} is more robust for multisource remote sensing registration.

Possible Errors and Model Limitations
Experimental results show that the SSSF descriptor is robust for multisource remote sensing image registration. Although some influencing factors, such as the template window sizes and canny operator thresholds are discussed, some possible problems may still arise. First, the proposed method is limited to image information and terrain structure. If the test data contain less scene shape information, the SSSF descriptor may achieve poor results in terms of matching accuracy. In hilly or mountainous regions, the terrain structure information of these areas is not obvious, because less contours and corners will be extracted. In addition, a large surface undulation causes significant local distortion. Second, the factor most affected by seasonality is vegetation. However, changes of vegetation usually do not affect the extraction of contours and corners, and the terrain structure information changes minimally. Therefore, seasonality has little effect on the method used in this study. Third, changes in small-scale land cover affect the extraction of contours and corners. Nevertheless, from a global perspective, extracting enough contours to construct SSSF descriptors and complete image matching is still possible. Seasonality and land cover usually have little effect on the proposed method, and the SSSF descriptors can always be built to match the tie points.

Discussion
This paper proposes a novel descriptor based on the shape context algorithm, SSSF descriptor. the normalized correlation coefficient of the SSSF descriptors (SSSF_{ncc}) as a similarity measure is applied to extract tie points with a template matching strategy. The piecewise linear transform model is then selected to rectify the remote sensing image. Compared with three state-of-the-art methods, the proposed one achieves the best matching accuracy because it integrates the terrain shape feature better and is robust to local rotation.

The canny operator is used to detect scene edges to construct the shape similarity in a local region. Different canny operator thresholds (i.e., from 0.1 to 0.3) were set to analyze the sensitivities of SSSF_{ncc} with regard to the changes of canny operator. Experimental results show that the proposed method obtains an improved matching outcome when the canny threshold is set to 0.2. The reason are as follows. The image contours detected by the canny threshold of 0.1 are too dense and plentiful to reduce the impact of the increase or decrease and deformation of terrain. By contrast, contours are not enough to use the canny threshold of 0.3 to extract the

terrain shape features, making it difficult to extract a sufficient number of tie points.

Although the proposed method has successfully used in multisource remote sensing image registration, some improvements to SSSF_{ncc} should be considered further. The different number of concentric circle layers and partitions can be set to analyze the influence of shape context algorithm for the SSSF descriptor, which changes the structure of the SSSF descriptor. Also, an image enhancement technology could be applied to enhance the shape and contour feature, which may be of some help to image matching. Only the SSSF descriptor of the template center is calculated, but SSSF_{ncc} is still a little time consuming compared with CSLTP since SSSF_{ncc} need to sample a number of contour points. In the future work, this issue could be resolved via a more efficient sampling method. As mentioned above, if the images contain less shape structure information, the SSSF_{ncc} may achieve poorer matching accuracy, because the SSSF_{ncc} depends on the local shape properties of images.

Conclusions
This study proposes a novel similarity measure named SSSF_{ncc} for multisource remote sensing registration. The metric is quite robust for multimodal remote sensing images with significant nonlinear grayscale differences. First, the shape context algorithm is used to extract SSSFs. Then, the scene shape similarity information is used to construct SSSF_{ncc}, and a template matching strategy is used to extract tie points. SSSF_{ncc} aims to acquire the scene shape similarity between multisource images and has been assessed by different categories of multisource test sets, including high-resolution, medium-resolution, and map sets. The registration results show that the SSSF_{ncc} is more robust than NCC and HOPC_{ncc} similarity measures for significant nonlinear grayscale differences. In summary, these registration results illustrate that the proposed method achieves robust registration outcomes.

References
Abdel-Hakim, A. E. and A. A. Farag. CSIFT: A SIFT descriptor with color invariant characteristics. Pages 1978–1983 in the 2006 IEEE Computer Society Conference on Proceedings of the Computer Vision and Pattern Recognition, 2006.

Belongie, S., J. Malik and J. Puzicha. 2001. Shape context: A new descriptor for shape matching and object recognition. In Advances in Neural Information Processing Systems 13, edited by T. K. Leen, T. G. Dietterich and V. Tresp, 831–837.

Brunner, D., G. Lemoine and L. Bruzzone. 2010. Earthquake damage assessment of buildings using VHR optical and SAR imagery. IEEE Transactions on Geoscience and Remote Sensing 48 (5):2403–2420.

Chen, S. H., X. R. Li, H. Yang and L. Y. Zhao. 2018. Robust local feature descriptor for multisource remote sensing image registration. Journal of Applied Remote Sensing 12.

Cole-Rhodes, A. A., K. L. Johnson, J. LeMoigne and I. Zavorin. 2003. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. IEEE Transactions on Image Processing 12 (12):1495–1511.

Fu, Z. T., Q. Q. Qin, B. Luo, C. Wu and H. Sun. 2019. A local feature descriptor based on combination of structure and texture information for multispectral image matching. IEEE Geoscience and Remote Sensing Letters 16 (1):100–104.

Goncalves, H., L. Corte-Real and J. A. Goncalves. 2011. Automatic image registration through image segmentation and SIFT. IEEE Transactions on Geoscience and Remote Sensing 49 (7):2589–2600.

Goncalves, H., J. A. Goncalves and L. Corte-Real. 2011. HAIRIS: A Method for automatic image registration through histogram-based image segmentation. IEEE Transactions on Image Processing 20 (3):776–789.

Hao, M., W. Z. Shi, H. Zhang C. and Li. 2014. Unsupervised change detection with expectation-maximization-based level set. IEEE Geoscience and Remote Sensing Letters 11 (1):210–214.

Hirschmuller, H. 2007. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2):328–341.

Hel-Or, Y., H. Hel-Or and E. David. 2014. Matching by tone mapping: Photometric invariant template matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (2):317–330.

Hong, G. and Y. Zhang. 2008. Wavelet-based image registration technique for high-resolution remote sensing images. Computers & Geosciences 34 (12):1708–1720.

Inglada, J. and A. Giros. 2004. On the possibility of automatic multisensor image registration. IEEE Transactions on Geoscience and Remote Sensing 42 (10):2104–2120.

Kelman, A., M. Sofka and C. V. Stewart. 2007. Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations. In 2007 IEEE Conference on Computer Vision and Pattern Recognition, DOI: 10.1109/CVPR.2007.383426.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60 (2):91–110.

Ma, J. L., J. C. W. Chan and F. Canters. 2010. Fully automatic subpixel image registration of multiangle CHRIS/Proba data. IEEE Transactions on Geoscience and Remote Sensing 48 (7):2829–2839.

Murphy, J. M., J. Le Moigne and D. J. Harding. 2016. Automatic image registration of multimodal remotely sensed data with global shearlet features. IEEE Transactions on Geoscience and Remote Sensing 54 (3):1685–1704.

Qu, X., S. Yue, G. Yue, Y. Shuang and L. Gao. 2016. A high-precision registration algorithm for heterologous image based on effective sub-graph extraction and feature points bidirectional matching. In Proceedings of the International Conference on Natural Computation.

Ravanbakhsh, M. and C. S. Fraser. 2013. A comparative study of DEM registration approaches. Journal of Spatial Science 58 (1):79–89.

Sedaghat, A. H. and H. Ebadi. 2015. Remote sensing image matching based on adaptive binning SIFT descriptor. IEEE Transactions on Geoscience and Remote Sensing 53 (10):5283–5293.

Sui, H. G., C. Xu, J. Y. Liu and F. Hua. 2015. Automatic optical-to-SAR image registration by iterative line extraction and Voronoi integrated spectral point matching. IEEE Transactions on Geoscience and Remote Sensing 53 (11):6058–6072.

Suri, S. and P. Reinartz. 2010. Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. IEEE Transactions on Geoscience and Remote Sensing 48 (2):939–949.

Xie, X., Y. Zhang, X. Ling and X. Wang. 2019. A novel extended phase correlation algorithm based on Log-Gabor filtering for multimodal remote sensing image registration. International Journal of Remote Sensing 40 (14):5429–5453.

Yan, L., Z. Q. Wang, Y. Liu and Z. Y. Ye. 2018. Generic and automatic Markov random field-based registration for multimodal remote sensing image using grayscale and gradient information. Remote Sensing 10 (8).

Ye, Y., J. Shan, L. Bruzzone and L. Shen. 2017. Robust registration of multimodal remote sensing images based on structural similarity. IEEE Transactions on Geoscience and Remote Sensing 55 (5):2941–2958.

Ye, Y. X. and J. Shan. 2014. A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences. ISPRS Journal of Photogrammetry and Remote Sensing 90:83–95.

Ye, Y. X., L. Shen, J. C. Wang, Z. P. Li and Z. Xu. 2015. Automatic matching of optical and SAR imagery through shape property. Pages 1072–1075 in 2015 IEEE International Geoscience and Remote Sensing Symposium.

Yu, L., D. R. Zhang and E. J. Holden. 2008. A fast and fully automatic registration approach based on point features for multi-source remote-sensing images. Computers & Geosciences 34 (7):838–848.

Zimmer, V. A., M. A. G. Ballester and G. Piella. 2019. Multimodal image registration using Laplacian commutators. Information Fusion 49:130–145.

Zitová, B. and J. Flusser. 2003. Image registration methods: A survey. Image & Vision Computing 21 (11):977–1000.